OCT 17 1989

1,282.00-101
07/810,560

501
Osteogenic Protein

Reference to Related Application

~~This application~~ which is a continuation-in-part of copending U.S. application Serial No. 315,342 filed February 23, 1989, which is a continuation-in-part of U.S. 232,630 filed August 15, 1988, which is a continuation-in-part of U.S. Serial No. 179,406 filed April 8, 1988, all entitled "Osteogenic Devices".

Background of the Invention

This invention relates to osteogenic protein, to genes encoding proteins which can induce osteogenesis in mammals and methods for their production using recombinant DNA techniques, and to bone and cartilage repair procedures using the osteogenic protein.

Mammalian bone tissue is known to contain one or more proteinaceous materials, presumably active during growth and natural bone healing, which can induce a developmental cascade of cellular events resulting in endochondral bone formation. This active factor (or factors) has variously been referred to in the literature as bone morphogenetic or morphogenic protein, bone inductive protein, osteogenic protein, osteogenin, or osteoinductive protein.

The developmental cascade of bone differentiation consists of recruitment of mesenchymal cells, proliferation of progenitor cells, calcification of cartilage, vascular invasion, bone formation, remodeling, and finally marrow differentiation (Reddi (1981) Collagen Rel. Res. 1:209-226).

Though the precise mechanisms underlying these phenotypic transformations are unclear, it has been shown that the natural endochondral bone differentiation activity of bone matrix can be dissociatively extracted and reconstituted with inactive residual collagenous matrix to restore full bone induction activity (Sampath and Reddi, (1981) Proc. Natl. Acad. Sci. USA 78:7599-7603). This provides an experimental method for assaying protein extracts for their ability to induce endochondral bone in vivo. Several species of mammals produce closely related protein as demonstrated by cross species implant experiments (Sampath and Reddi (1983) Proc. Natl. Acad. Sci. USA 80:6591-6595).

The potential utility of these proteins has been recognized widely. It is contemplated that the availability of the protein would revolutionize orthopedic medicine, certain types of plastic surgery, and various periodontal and craniofacial reconstructive procedures.

The observed properties of these protein fractions have induced an intense research effort in various laboratories directed to isolating and identifying the pure factor or factors responsible for osteogenic activity. The current state of the art of purification of osteogenic protein from mammalian bone is disclosed by Sampath et al. (Proc. Natl. Acad. Sci. USA (1987) 80). Urist et al. (Proc. Soc. Exp. Biol. Med. (1984) 173:194-199) disclose a human

osteogenic protein fraction which was extracted from demineralized cortical bone by means of a calcium chloride-urea inorganic-organic solvent mixture, and retrieved by differential precipitation in guanidine-hydrochloride and preparative gel electrophoresis. The authors report that the protein fraction has an amino acid composition of an acidic polypeptide and a molecular weight in a range of 17-18 kD.

Urist et al. (Proc. Natl. Acad. Sci. USA (1984) 81:371-375) disclose a bovine bone morphogenetic protein extract having the properties of an acidic polypeptide and a molecular weight of approximately 18 kD. The authors reported that the protein was present in a fraction separated by hydroxyapatite chromatography, and that it induced bone formation in mouse hindquarter muscle and bone regeneration in trephine defects in rat and dog skulls. Their method of obtaining the extract from bone results in ill-defined and impure preparations.

European Patent Application Serial No. 148,155, published October 7, 1985, purports to disclose osteogenic proteins derived from bovine, porcine, and human origin. One of the proteins, designated by the inventors as a P3 protein having a molecular weight of 22-24 kD, is said to have been purified to an essentially homogeneous state. This material is reported to induce bone formation when implanted into animals.

International Application No. PCT/087/01537, published January 14, 1988, discloses an impure fraction from bovine bone which has bone induction qualities. The named applicants also disclose putative "bone inductive factors" produced by recombinant DNA techniques. Four DNA sequences were retrieved from human or bovine genomic or cDNA libraries

and expressed in recombinant host cells. While the applicants stated that the expressed proteins may be bone morphogenic proteins, bone induction was not demonstrated, suggesting that the recombinant proteins are not osteogenic. The same group reported subsequently (Science, V. 242, p. 1528, Dec, 1988) that three of the four factors induce cartilage formation, and postulate that bone formation activity "is due to a mixture of regulatory molecules" and that "bone formation is most likely controlled ... by the interaction of these molecules." Again, no bone induction was attributed to the products of expression of the cDNAs. See also Urist et al., EP 0,212,474 entitled Bone Morphogenic Agents.

lid 4-90 NMW

Wang et al. (Proc. Nat. Acad. Sci. USA (1988) 85: 9484-9488) discloses the purification of a bovine bone morphogenetic protein from guanidine extracts of demineralized bone having cartilage and bone formation activity as a basic protein corresponding to a molecular weight of 30 kD determined from gel elution. Purification of the protein yielded proteins of 30, 18 and 16 kD which, upon separation, were inactive. In view of this result, the authors acknowledged that the exact identity of the active material had not been determined.

It is an object of this invention to provide osteogenic devices comprising matrices containing dispersed osteogenic protein capable of bone induction in allogenic and xenogenic implants. Another object is to characterize the protein responsible for osteogenesis. Another object is to provide natural and recombinant osteogenic proteins capable of inducing endochondral bone formation in mammals, including humans. Yet another object is to provide genes encoding osteogenic proteins and methods for their production using recombinant DNA techniques.

These and other objects and features of the invention will be apparent from the description, drawings, and claims which follow.

Summary of the Invention

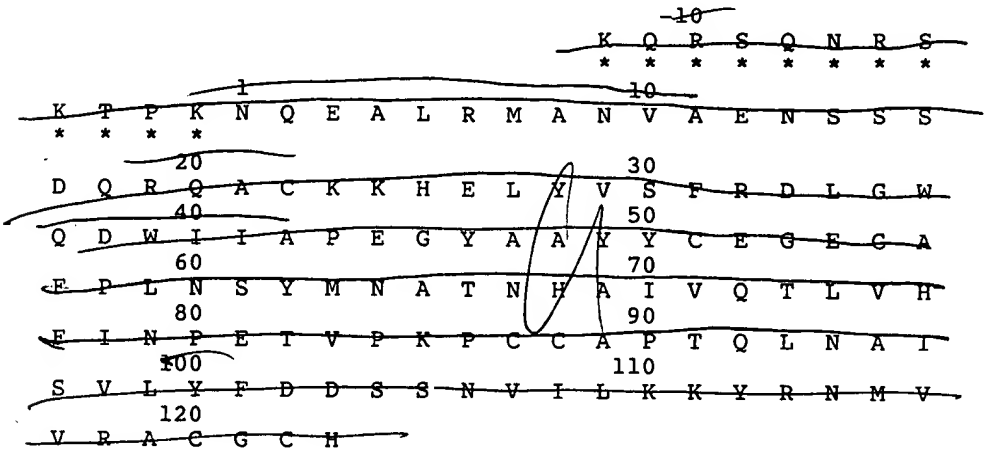
This invention provides osteogenic proteins and devices which, when implanted in a mammalian body, can induce at the locus of the implant the full developmental cascade of endochondral bone formation and bone marrow differentiation. The devices comprise a carrier material, referred to herein as a matrix, having the characteristics disclosed below, containing dispersed osteogenic protein produced using recombinant DNA techniques.

A key to these developments was the elucidation of amino acid sequence and structure data of native osteogenic protein. A protocol was developed which results in retrieval of active, substantially pure osteogenic protein from mammalian bone having a half-maximum bone forming activity of about 0.8 to 1.0 ng per mg of implant. The availability of the material enabled the inventors to elucidate all structural details of the protein necessary to achieve bone formation. Knowledge of the protein's amino acid sequence and other structural features enabled the identification and cloning of native genes.

Consensus DNA sequences based on partial sequence data and observed homologies with regulatory proteins disclosed in the literature were used as probes for extracting genes encoding osteogenic protein from genomic and cDNA libraries. One of the consensus sequences was used to isolate a heretofore unidentified DNA sequence, portions of which, when ligated, encode a protein comprising a region capable of inducing endochondral bone formation when properly modified, incorporated in a suitable matrix, and implanted as disclosed herein. Messenger RNA corresponding to the DNA was retrieved from human cells.

The sequence of what is believed to be the mature form of the new protein, designated OP1, is:

OP1



There remains uncertainty as to the N-terminus of the native form. It is believed that maturation by cleavage occurs somewhere in the 11 to 21 amino acid region. The full length protein, expressed from the cDNA, has the sequence set forth in Fig. 1B.

The active region (functional domain) of this protein has the amino acid sequence:

OPS

```
                                    30
                                L   Y   V   S   F   R   D   L   G   W
            40                                  50
Q   D   W   I   I   A   P   E   G   Y   A   A   Y   Y   C   E   G   E   C   A
            60                                  70
F   P   L   N   S   Y   M   N   A   T   N   H   A   I   V   Q   T   L   V   H
            80                                  90
F   I   N   P   E   T   V   P   K   P   C   C   A   P   T   Q   L   N   A   I
           100                                 110
S   V   L   Y   F   D   D   S   S   N   V   I   L   K   K   Y   R   N   M   V
           120
V   R   A   C   G   C   H
```

A longer active sequence is:

OP7

```
                                C   K   K   H   E   L   Y   V   30
                                                                S   F   R   D   L   G   W
            40                                                  50
Q   D   W   I   I   A   P   E   G   Y   A   A   Y   Y   C   E   G   E   C   A
            60                                                  70
F   P   L   N   S   Y   M   N   A   T   N   H   A   I   V   Q   T   L   V   H
            80                                                  90
F   I   N   P   E   T   V   P   K   P   C   C   A   P   T   Q   L   N   A   I
           100                                                 110
S   V   L   Y   F   D   D   S   S   N   V   I   L   K   K   Y   R   N   M   V
           120
V   R   A   C   G   C   H   *
```

Fig. 1A discloses the genomic DNA sequence of OP1.

The invention provides recombinant dimeric proteins, and osteogenic devices comprising the proteins, comprising a pair of disulfide bonded subunits, at least one of which includes the active region of OP1, identified above.

The probe also retrieved DNA sequences identified in PCT/087/01537, referenced above, designated therein as BMP2(b). The inventors herein discovered that certain

subparts of the designated BMP2(a) and BMP2(b) sequences, when properly assembled, encode proteins (CBMP2a and CBMP2b) which have true osteogenic activity, i.e., induce the full cascade of events leading to endochondral bone formation when properly folded, dimerized, and implanted in a mammal. The active regions of these sequences are:

CBMP2AS

```
                                                                L  Y
                        30                                          40
V  D  F  S  D  V  G  W  N  D  W  I  V  A  P  P  G  Y  H  A
                     50                                          60
F  Y  C  H  G  E  C  P  F  P  L  A  D  H  L  N  S  T  N  H
                     70                                          80
A  I  V  Q  T  L  V  N  S  V  N  S  K  I  P  K  A  C  C  V
                     90                                         100
P  T  E  L  S  A  I  S  M  L  Y  L  D  E  N  E  K  V  V  L
                    110
K  N  Y  Q  D  M  V  V  E  G  C  G  C  R
```

and

CBMP2BS

```
                                 30                                 40
      L  Y  V  D  F  S  D  V  G  W  N  D  W  I  V  A  P
                              50                                    60
P  G  Y  Q  A  F  Y  C  H  G  D  C  P  F  P  L  A  D  H  L
                              70                                    80
N  S  T  N  H  A  I  V  Q  T  L  V  N  S  V  N  S  S  I  P
                              90                                   100
K  A  C  C  V  P  T  E  L  S  A  I  S  M  L  Y  L  D  E  Y
                             110                                  120
D  K  V  V  L  K  N  Y  Q  E  M  V  V  E  G  C  G  C  R  *
```

Longer active regions are:

CBMP2AL

```
                                                20
                                    C  K  R  H  P  L  Y
                     30                         40
V  D  F  S  D  V  G  W  N  D  W  I  V  A  P  P  G  Y  H  A
                     50                         60
F  Y  C  H  G  E  C  P  F  P  L  A  D  H  L  N  S  T  N  H
                     70                         80
A  I  V  Q  T  L  V  N  S  V  N  S  K  I  P  K  A  C  C  V
                     90                        100
P  T  E  L  S  A  I  S  M  L  Y  L  D  E  N  E  K  V  V  L
                    110
K  N  Y  Q  D  M  V  V  E  G  C  G  C  R  *
```

and

CBMP2BL

```
                                                         20
                                                      C  R
                           30                            40
R  H  S  L  Y  V  D  F  S  D  V  G  W  N  D  W  I  V  A  P
                           50                            60
P  G  Y  Q  A  F  Y  C  H  G  D  C  P  F  P  L  A  D  H  L
                           70                            80
N  S  T  N  H  A  I  V  Q  T  L  V  N  S  V  N  S  S  I  P
                           90                           100
K  A  C  C  V  P  T  E  L  S  A  I  S  M  L  Y  L  D  E  Y
                          110                           120
D  K  V  V  L  K  N  Y  Q  E  M  V  V  E  G  C  G  C  R  *
```

What is believed to be the mature forms of these proteins is:

CBMP2AM

```
                                -10                                 -1
                        ........G  K  G  H  P  L  H  K  R  E  K  R
                                10                                  20
Q  A  K  H  K  Q  R  K  R  L  K  S  S  C  K  R  H  P  L  Y
                                30                                  40
V  D  F  S  D  V  G  W  N  D  W  I  V  A  P  P  G  Y  H  A
                                50                                  60
F  Y  C  H  G  E  C  P  F  P  L  A  D  H  L  N  S  T  N  H
                                70                                  80
A  I  V  Q  T  L  V  N  S  V  N  S  K  I  P  K  A  C  C  V
                                90                                 100
P  T  E  L  S  A  I  S  M  L  Y  L  D  E  N  E  K  V  V  L
                               110
K  N  Y  Q  D  M  V  V  E  G  C  G  C  R  *
```

and

CBMP2BM

```
                                -10                                 -1
                        ........R  G  H  A  L  T  R  R  R  R
                                10                                  20
A  K  R  S  P  K  H  H  S  Q  R  A  R  K  K  N  K  N  C  R
                                30                                  40
R  H  S  L  Y  V  D  F  S  D  V  G  W  N  D  W  I  V  A  P
                                50                                  60
P  G  Y  Q  A  F  Y  C  H  G  D  C  P  F  P  L  A  D  H  L
                                70                                  80
N  S  T  N  H  A  I  V  Q  T  L  V  N  S  V  N  S  S  I  P
                                90                                 100
K  A  C  C  V  P  T  E  L  S  A  I  S  M  L  Y  L  D  E  Y
                               110                                 120
D  K  V  V  L  K  N  Y  Q  E  M  V  V  E  G  C  G  C  R  *
```

There remains uncertainty as to the N-terminus of the native form of these proteins. It is believed that maturation by cleavage occurs somewhere in the -11 to -1 positions. Figures 2 and 3 disclose the full length protein expressed from the human cDNA of CBMP2a and CBMP2b.

The osteogenic protein in its native form is a glycosylated dimer and has an apparent molecular weight of about 30 kD as determined by SDS-PAGE. When reduced, the 30 kD protein gives rise to two glycosylated polypeptide chains (subunits) having apparent molecular weights of about 16 kD and 18 kD. In the reduced state, the 30 kD protein has no detectable osteogenic activity. The deglycosylated dimer, which has osteogenic activity, has an apparent molecular weight of about 27 kD. When reduced, the 27 kD protein separates into two deglycosylated subunits having molecular weights of about 14 kD and 16 kD. The larger of these subunits is mature OP1, the other is mature CBMP2(a) or CBMP2(b) CBMP2(b) differs from 2(a) at only five residues in the active region. Human DNA sequence data suggest that the second subunit is CBMP2(b). Bovine protein sequence data suggest that the second subunit is CBMP2(a). Recombinant versions of both are active cross species.

Various of these proteins have been expressed from intact or truncated cDNA or from synthetic DNAs in E. coli as fusion proteins, and then purified, cleaved, refolded, dimerized, and implanted in experimental animals. The recombinant constructs induce the full cascade of endochondral bone formation. Homodimers of CBMP2(a) and (b) are osteoinductive. Higher activity apparently is achieved with homodimers of OP1. Heterodimers of OP1 and BMP2 also are active. The osteoinductive effect is not dependent on the presence of the entire mature form amino acid sequences of either subunit. Properly folded dimers comprising minimal structure, as short as 94 amino acids, are active. Furthermore, analogs of the active region are active (see, e.g., the COP5 and COP7 constructs, ~~supra~~ infra). Thus, the osteogenic protein of the invention may include forms having varying glycosylation patterns, varying N-termini, a family of

related proteins having regions of amino acid sequence homology, and active truncated or mutated forms of native protein, produced by expression of recombinant DNA in host cells.

Thus, in view of this disclosure, skilled genetic engineers can isolate genes from cDNA or genomic libraries which encode appropriate amino acid sequences, or construct DNAs from oligonucleotides, and then can express them in various types of host cells, including both procaryotes and eucaryotes, to produce large quantities of active proteins capable of inducing bone formation in mammals including humans.

The osteogenic proteins are useful in clinical applications in conjunction with a suitable delivery or support system (matrix). The matrix is made up of particles or porous materials. The pores must be of a dimension to permit progenitor cell migration and subsequent differentiation and proliferation. The particle size should be within the range of 70 - 850 μm, preferably 150μm - 420μm. It may be fabricated by close packing particulate material into a shape spanning the bone defect, or by otherwise structuring as desired a material that is biocompatible (non-inflammatory) and, biodegradable in vivo to serve as a "temporary scaffold" and substratum for recruitment of migratory progenitor cells, and as a base for their subsequent anchoring and proliferation. Currently preferred carriers include particulate, demineralized, guanidine extracted, species-specific (allogenic) bone, and specially treated particulate, protein extracted, demineralized, xenogenic bone. Optionally, such xenogenic bone powder matrices also may be treated with proteases such as trypsin, swelling agents such as dichloromethane, trichloroacetic acid, and acetonitrile and acids such as trifluoroacetic acid and

hydrogen fluoride. Other potentially useful matrix materials comprise collagen, homopolymers and copolymers of glycolic acid and lactic acid, hydroxyapatite, tricalcium phosphate and other calcium phosphates.

The osteogenic proteins and implantable osteogenic devices enabled and disclosed herein will permit the physician to obtain optimal predictable bone formation to correct, for example, acquired and congenital craniofacial and other skeletal or dental anomalies (Glowacki et al. (1981) Lancet 1:959-963). The devices may be used to induce local endochondral bone formation in non-union fractures as demonstrated in animal tests, and in other clinical applications including periodontal applications where bone formation is required. Another potential clinical application is in cartilage repair, for example, in the treatment of osteoarthritis.

Brief Description of the Drawing

The foregoing and other objects of this invention, the various features thereof, as well as the invention itself, may be more fully understood from the following description, when read together with the accompanying drawings, in which:

FIGURE 1A represents the genomic nucleotide sequence of OP1;

FIGURE 1B represents the full length cDNA and encoded amino acid sequence of the prepro form of human OP1 protein;

FIGURE 1C is a representation of the hybridization of the consensus gene/probe to the OP1 gene;

FIGURE 2 represents the full length cDNA and encoded amino acid sequence of the prepro form of human CBMP2(a); and

FIGURE 3 represents the full length cDNA and encoded amino acid sequence of the prepro form of human CBMP2(b).

Description

Purification protocols were developed which enabled isolation of the osteogenic protein present in crude protein extracts from mammalian bone. (See PCT US 89/01453, and U.S. Serial No. 179,406 filed April 8, 1988). The development of the procedure, coupled with the availability of fresh calf bone, enabled isolation of substantially pure bovine osteogenic protein (BOP). BOP was characterized significantly; its ability to induce cartilage and ultimately endochondral bone growth in cat, rabbit, and rat were demonstrated and studied; it was shown to be able to induce the full developmental cascade of bone formation previously ascribed to unknown protein or proteins in heterogeneous bone extracts. Sequence data obtained from the bovine materials suggested probe designs which were used to isolate human genes. The BOP human counterpart proteins have now been expressed and extensively characterized.

These discoveries enable preparation of DNAs encoding totally novel, non-native protein constructs which individually as homodimers and combined with other species as heterodimers are capable of producing true endochondral bone. They also permit expression of the natural material, truncated forms, muteins, analogs, fusion proteins, and various other variants and constructs, from cDNAs and genomic DNAs retrieved from natural sources or from synthetic DNA produced using the techniques disclosed herein and automated, commercially available equipment. The DNAs may be expressed using well established recombinant DNA technologies in procaryotic or eucaryotic host cells, and may be oxidized and refolded in vitro if necessary for biological activity.

More specifically, a synthetic consensus gene shown in FIGURE 1C, was designed as a hybridization probe. The design was based on amino acid sequence data obtained by sequencing digestion fragments of naturally sourced material and on predictions from observed homologies of these sequences with the TGF-beta gene family. The consensus gene/probe exploited human codon bias as found in human TGF-beta. The designed sequence was then constructed using known techniques of assembly of oligonucleotides manufactured in a DNA synthesizer.

A human genomic library (Maniatis-library) carried in lambda phage (Charon 4A) was screened using the probe. Twenty-four positive clones were found. Five contained a gene never before reported designated OP1, osteogenic protein-1, described below. Two others yielded genes corresponding to "BMP-2b", (see PCT US 87/01537).

Southern blot analysis of lambda #13 DNA showed that an approximately 3kb BamHI fragment hybridized to the probe. (See Fig. 1C). This fragment was isolated and subcloned. Analysis of this sequence showed that the fragment encoded the carboxy terminus of a protein, named OP1. The protein was identified by amino acid homology with the TGF-beta family. Consensus splice signals were found where amino acid homologies ended, designating exon-intron boundaries. Three exons were combined to obtain a functional TGF-beta-like domain containing seven cysteines. The DNA sequence of the functional domain was then used as a probe to screen a human cDNA library as described below. The probe also retrieved sequences of BMP2(b).

The OP1 and BMP2 sequences labeled with 32P were used to probe a human placenta 5' strech lambda phage cDNA library (Clonetech, Palo Alto, California), and a human hippocampus library (Strategene, Inc.). Positive sequences hybridized at low stringency enabled isolation of full length cDNAs (translated region) of both CBMP2(b) and OPI. These cDNA sequences, and the prepro form amino acid sequences they encode, are set forth in Fig. 3 and Fig. 1B.

The amino acid sequence of the OPI active region is set forth below:

OPS

```
                                        30
                                    L   Y   V   S   F   R   D   L   G   W
            40                                  50
Q   D   W   I   I   A   P   E   G   Y   A   A   Y   Y   C   E   G   E   C   A
            60                                  70
F   P   L   N   S   Y   M   N   A   T   N   H   A   I   V   Q   T   L   V   H
            80                                  90
F   I   N   P   E   T   V   P   K   P   C   C   A   P   T   Q   L   N   A   I
          100                                 110
S   V   L   Y   F   D   D   S   S   N   V   I   L   K   K   Y   R   N   M   V
          120
V   R   A   C   G   C   H
```

A longer active sequence is:

OP7

```
                                        30
                    C   K   K   H   E   L   Y   V   S   F   R   D   L   G   W
            40                                  50
Q   D   W   I   I   A   P   E   G   Y   A   A   Y   Y   C   E   G   E   C   A
            60                                  70
F   P   L   N   S   Y   M   N   A   T   N   H   A   I   V   Q   T   L   V   H
            80                                  90
F   I   N   P   E   T   V   P   K   P   C   C   A   P   T   Q   L   N   A   I
          100                                 110
S   V   L   Y   F   D   D   S   S   N   V   I   L   K   K   Y   R   N   M   V
          120
V   R   A   C   G   C   H   *
```

Another active sequence is

~~The mature form has the structure~~:

OPM

```
                                                                -10
                                                          K  Q  R  S  Q  N  R  S
                                                          *  *  *  *  *  *  *  *
                    1                                          10
K  T  P  K  N  Q  E  A  L  R  M  A  N  V  A  E  N  S  S  S
*  *  *  *                                                    20                                          30
D  Q  R  Q  A  C  K  K  H  E  L  Y  V  S  F  R  D  L  G  W
          40                                          50
Q  D  W  I  I  A  P  E  G  Y  A  A  Y  Y  C  E  G  E  C  A
          60                                          70
F  P  L  N  S  Y  M  N  A  T  N  H  A  I  V  Q  T  L  V  H
          80                                          90
F  I  N  P  E  T  V  P  K  P  C  C  A  P  T  Q  L  N  A  I
         100                                         110
S  V  L  Y  F  D  D  S  S  N  V  I  L  K  K  Y  R  N  M  V
         120
V  R  A  C  G  C  H
```

~~where residue 1 is the presumed N terminus~~.

Referring to Figure 1B, the amino acid sequence of the prepro form of OP1 is depicted. The underlined residues are believed to constitute a cleavage site where the N terminal residues, constituting a signal peptide, are removed as the remaining pro form is secreted from the expressing cell. The starred sequence represents the region where the pro form is believed to be cut to produce the mature form. Cleavage somewhere about the starred residues is hypothesized based on the amino acid sequence similarity in that region with TGFß superfamily proteins and what is known about their posttranslational modification, and also on molecular weight (electrophoretic migration) data obtained from natural sourced, active BOP. Both the pro form and prepro form, when properly dimerized, folded, adsorbed on a matrix, and implanted, display osteogenic activity, presumably due to

proteolytic degradation resulting in cleavage and generation of mature form protein or active truncated analogs.

The amino acid sequence of the active region (functional domain) encoded by the CBMP2(b) native gene retrieved using the consensus probe is:

CBMP2BS

```
                                30                                  40
        L  Y  V  D  F  S  D  V  G  W  N  D  W  I  V  A  P
                                50                                  60
P  G  Y  Q  A  F  Y  C  H  G  D  C  P  F  P  L  A  D  H  L
                                70                                  80
N  S  T  N  H  A  I  V  Q  T  L  V  N  S  V  N  S  S  I  P
                                90                                  100
K  A  C  C  V  P  T  E  L  S  A  I  S  M  L  Y  L  D  E  Y
                               110                                  120
D  K  V  V  L  K  N  Y  Q  E  M  V  V  E  G  C  G  C  R  *
```

The amino acid sequence of the active region of CBMP2a is:

CBMP2AS

```
                                                                L  Y
                        30                                  40
V  D  F  S  D  V  G  W  N  D  W  I  V  A  P  P  G  Y  H  A
                        50                                  60
F  Y  C  H  G  E  C  P  F  P  L  A  D  H  L  N  S  T  N  H
                        70                                  80
A  I  V  Q  T  L  V  N  S  V  N  S  K  I  P  K  A  C  C  V
                        90                                  100
P  T  E  L  S  A  I  S  M  L  Y  L  D  E  N  E  K  V  V  L
                       110
K  N  Y  Q  D  M  V  V  E  G  C  G  C  R
```

Longer active regions are:

CBMP2BL

```
                                                                  20
                                                              C   R
                                          30                      40
R  H  S  L  Y  V  D  F  S  D  V  G  W  N  D  W  I  V  A  P
                                          50                      60
P  G  Y  Q  A  F  Y  C  H  G  D  C  P  F  P  L  A  D  H  L
                                          70                      80
N  S  T  N  H  A  I  V  Q  T  L  V  N  S  V  N  S  S  I  P
                                          90                     100
K  A  C  C  V  P  T  E  L  S  A  I  S  M  L  Y  L  D  E  Y
                                         110                     120
D  K  V  V  L  K  N  Y  Q  E  M  V  V  E  G  C  G  C  R  *
```

CBMP2AL

```
                                                        20
                                                  C  K  R  H  P  L  Y
                                 30                        40
V  D  F  S  D  V  G  W  N  D  W  I  V  A  P  P  G  Y  H  A
                                 50                        60
F  Y  C  H  G  E  C  P  F  P  L  A  D  H  L  N  S  T  N  H
                                 70                        80
A  I  V  Q  T  L  V  N  S  V  N  S  K  I  P  K  A  C  C  V
                                 90                       100
P  T  E  L  S  A  I  S  M  L  Y  L  D  E  N  E  K  V  V  L
                                110
K  N  Y  Q  D  M  V  V  E  G  C  G  C  R  *
```

The mature forms of these proteins have the following sequence (presumed N terminal):

CBMP2BM

```
                                    -10                                 -1
                          ........R  G  H  A  L  T  R  R  R  R
                                10                                      20
A  K  R  S  P  K  H  H  S  Q  R  A  R  K  K  N  K  N  C  R
                                30                                      40
R  H  S  L  Y  V  D  F  S  D  V  G  W  N  D  W  I  V  A  P
                                50                                      60
P  G  Y  Q  A  F  Y  C  H  G  D  C  P  F  P  L  A  D  H  L
                                70                                      80
N  S  T  N  H  A  I  V  Q  T  L  V  N  S  V  N  S  S  I  P
                                90                                     100
K  A  C  C  V  P  T  E  L  S  A  I  S  M  L  Y  L  D  E  Y
                               110                                     120
D  K  V  V  L  K  N  Y  Q  E  M  V  V  E  G  C  G  C  R  *
```

and

CBMP2AM

```
                          -10                                 -1
                    ........G  K  G  H  P  L  H  K  R  E  K  R
                          10                                      20
Q  A  K  H  K  Q  R  K  R  L  K  S  S  C  K  R  H  P  L  Y
                          30                                      40
V  D  F  S  D  V  G  W  N  D  W  I  V  A  P  P  G  Y  H  A
                          50                                      60
F  Y  C  H  G  E  C  P  F  P  L  A  D  H  L  N  S  T  N  H
                          70                                      80
A  I  V  Q  T  L  V  N  S  V  N  S  K  I  P  K  A  C  C  V
                          90                                     100
P  T  E  L  S  A  I  S  M  L  Y  L  D  E  N  E  K  V  V  L
                         110
K  N  Y  Q  D  M  V  V  E  G  C  G  C  R  *
```

Again, there is uncertainty regarding where cleavage occurs.

Given the foregoing amino acid and DNA sequence information, various DNAs can be constructed which encode at least the minimal active domain of OP1 and/or BMP2, various analogs thereof, as well as fusion proteins, truncated forms of the mature proteins, and similar constructs. The pro form

and the prepro form are active, presumably because of in situ cleavage events or generation of active products by cleavage during protein processing. These DNAs can be produced by those skilled in the art using well known DNA manipulative techniques involving genomic and cDNA isolation, construction of synthetic DNA from synthesized oligonucleotides, and cassette mutagenesis techniques. 15-100mer oligonucleotides may be synthesized on a Biosearch DNA Model 8600 Synthesizer, and purified by polyacrylamide gel electrophoresis (PAGE) in Tris-Borate-EDTA buffer. The DNA is then electroeluted from the gel. Overlapping oligomers may be phosphorylated by T4 polynucleotide kinase and ligated into larger blocks which may also be purifed by PAGE.

The cDNA or synthetic DNA may be integrated into an expression vector and transfected into an appropriate host cell for protein expression. The host may be a procaryotic or eucaryotic cell since the former's inability to glycosylate protein will not destroy the protein's osteogenic activity. Useful host cells include Saccharomyces, E. coli, and various mammalian cell cultures. The vector may additionally encode various signal sequences for protein secretion and/or may encode osteogenic protein as a fusion protein. After being translated, protein may be purified from the cells or recovered from the culture medium. All active protein forms comprise dimeric species joined by disulfide bonds or otherwise associated, produced by oxidizing and refolding one or more of the various recombinant proteins within an appropriate eucaryotic cell or in vitro after expression of individual subunits.

Using such techniques, the following DNAs were prepared:

OP1 Fusion Proteins

OP1A

```
        10         20         30         40         50         60
ATGAAAGCAATTTTCGTACTGAAAGGTTCACTGGACAGAGATCTGGACTCTCGTCTGGAT
 M  K  A  I  F  V  L  K  G  S  L  D  R  D  L  D  S  R  L  D
        70         80         90        100        110        120
CTGGACGTTCGTACCGACCACAAAGACCTGTCTGATCACCTGGTTCTGGTCGACCTGGCT
 L  D  V  R  T  D  H  K  D  L  S  D  H  L  V  L  V  D  L  A
       130        140        150        160        170        180
CGTAACGACCTGGCTCGTATCGTTACTCCCGGGTCTCGTTACGTTGCGGATCTGGAATTC
 R  N  D  L  A  R  I  V  T  P  G  S  R  Y  V  A  D  L  E  F
       190        200        210        220        230        240
GATCCTCACCAGAGGCAGGCCTGTAAGAAGCACGAGCTGTATGTCAGCTTCCGAGACCTG
 D  P  H  Q  R  Q  A  C  K  K  H  E  L  Y  V  S  F  R  D  L
 *  *
       250        260        270        280        290        300
GGCTGGCAGGACTGGATCATCGCGCCTGAAGGCTACGCGGCCTACTACTGTGAGGGGGAG
 G  W  Q  D  W  I  I  A  P  E  G  Y  A  A  Y  Y  C  E  G  E
       310        320        330        340        350        360
TGTGCCTTCCCTCTGAACTCCTACATGAACGCCACCAACCACGCCATCGTGCAGACGCTG
 C  A  F  P  L  N  S  Y  M  N  A  T  N  H  A  I  V  Q  T  L
       370        380        390        400        410        420
GTCCACTTCATCAACCCGGAAACGGTGCCCAAGCCCTGCTGTGCGCCCACGCAGCTCAAT
 V  H  F  I  N  P  E  T  V  P  K  P  C  C  A  P  T  Q  L  N
       430        440        450        460        470        480
GCCATCTCCGTCCTCTACTTCGATGACAGCTCCAACGTCATCCTGAAGAAATACAGAAAC
 A  I  S  V  L  Y  F  D  D  S  S  N  V  I  L  K  K  Y  R  N
       490        500        510
ATGGTGGTCCGGGCCTGTGGCTGCCACTAACTGCAG
 M  V  V  R  A  C  G  C  H
```

OP1B

```
        10         20         30         40         50         60
ATGAAAGCAATTTTCGTACTGAAAGGTTCACTGGACAGAGATCTGGACTCTCGTCTGGAT
 M  K  A  I  F  V  L  K  G  S  L  D  R  D  L  D  S  R  L  D
        70         80         90        100        110        120
CTGGACGTTCGTACCGACCACAAAGACCTGTCTGATCACCTGGTTCTGGTCGACCTGGCT
 L  D  V  R  T  D  H  K  D  L  S  D  H  L  V  L  V  D  L  A
                                                   op-1|----->
       130        140        150        160        170        180
CGTAACGACCTGGCTCGTATCGTTACTCCCGGGTCTCGTTACGTTGCGGATCTGGAATTC
 R  N  D  L  A  R  I  V  T  P  G  S  R  Y  V  A  D  L  E  F
```

```
       190        200        210        220        230        240
CGGATCTACAAGGACTACATCCGGGAACGCTTCGACAATGAGACGTTCCGGATCAGCGTT
 R  I  Y  K  D  Y  I  R  E  R  F  D  N  E  T  F  R  I  S  V
       250        260        270        280        290        300
TATCAGGTGCTCCAGGAGCACTTGGGCAGGGAATCGGATCTCTTCCTGCTCGACAGCCGT
 Y  Q  V  L  Q  E  H  L  G  R  E  S  D  L  F  L  L  D  S  R
       310        320        330        340        350        360
ACCCTCTGGGCCTCGGAGGAGGGCTGGCTGGTGTTTGACATCACAGCCACCAGCAACCAC
 T  L  W  A  S  E  E  G  W  L  V  F  D  I  T  A  T  S  N  H
       370        380        390        400        410        420
TGGGTGGTCAATCCGCGGCACAACCTGGGCCTGCAGCTCTCGGTGGAGACGCTGGATGGG
 W  V  V  N  P  R  H  N  L  G  L  Q  L  S  V  E  T  L  D  G
       430        440        450        460        470        480
CAGAGCATCAACCCCAAGTTGGCGGGCCTGATTGGGCGGCACGGGCCCCAGAACAAGCAG
 Q  S  I  N  P  K  L  A  G  L  I  G  R  H  G  P  Q  N  K  Q
       490        500        510        520        530        540
CCCTTCATGGTGGCTTTCTTCAAGGCCACGGAGGTCCACTTCCGCAGCATCCGGTCCACG
 P  F  M  V  A  F  F  K  A  T  E  V  H  F  R  S  I  R  S  T
       550        560        570        580        590        600
GGGAGCAAACAGCGCAGCCAGAACCGCTCCAAGACGCCCAAGAACCAGGAAGCCCTGCGG
 G  S  K  Q  R  S  Q  N  R  S  K  T  P  K  N  Q  E  A  L  R
       610        620        630        640        650        660
ATGGCCAACGTGGCAGAGAACAGCAGCAGCGACCAGAGGCAGGCCTGTAAGAAGCACGAG
 M  A  N  V  A  E  N  S  S  S  D  Q  R  Q  A  C  K  K  H  E
       670        680        690        700        710        720
CTGTATGTCAGCTTCCGAGACCTGGGCTGGCAGGACTGGATCATCGCGCCTGAAGGCTAC
 L  Y  V  S  F  R  D  L  G  W  Q  D  W  I  I  A  P  E  G  Y
       730        740        750        760        770        780
GCCGCCTACTACTGTGAGGGGGAGTGTGCCTTCCCTCTGAACTCCTACATGAACGCCACC
 A  A  Y  Y  C  E  G  E  C  A  F  P  L  N  S  Y  M  N  A  T
       790        800        810        820        830        840
AACCACGCCATCGTGCAGACGCTGGTCCACTTCATCAACCCGGAAACGGTGCCCAAGCCC
 N  H  A  I  V  Q  T  L  V  H  F  I  N  P  E  T  V  P  K  P
       850        860        870        880        890        900
TGCTGTGCGCCCACGCAGCTCAATGCCATCTCCGTCCTCTACTTCGATGACAGCTCCAAC
 C  C  A  P  T  Q  L  N  A  I  S  V  L  Y  F  D  D  S  S  N
       910        920        930        940        950        960
GTCATCCTGAAGAAATACAGAAACATGGTGGTCCGGGCCTGTGGCTGCCACTAGCTCCTT
 V  I  L  K  K  Y  R  N  M  V  V  R  A  C  G  C  H  *
       970        980        990       1000
CCGAGAATTCCAGACCTTTGGGGCCCAAAGGTTTTTCTGGATCC
```

OP1C

```
        10         20         30         40         50         60
ATGAAAGCAATTTTCGTACTGAAAGGTTCACTGGACAGAGATCTGGACTCTCGTCTGGAT
 M  K  A  I  F  V  L  K  G  S  L  D  R  D  L  D  S  R  L  D
        70         80         90        100        110        120
CTGGACGTTCGTACCGACCACAAAGACCTGTCTGATCACCTGGTTCTGGTCGACCTGGCT
 L  D  V  R  T  D  H  K  D  L  S  D  H  L  V  L  V  D  L  A
       130        140        150        160        170        180
CGTAACGAGAATTCCCGGGTAGCGCGTAGAGCCGGCGCGATGCACGTGCGCTCACTGCGA
 R  N  E  N  S  R  V  A  R  R  A  G  A  M  H  V  R  S  L  R
                                           |-----------opl-----
       190        200        210        220        230        240
GCTGCGGCGCCGCACAGCTTCGTGGCGCTCTGGGCACCCCTGTTCCTGCTGCGCTCCGCC
 A  A  A  P  H  S  F  V  A  L  W  A  P  L  F  L  L  R  S  A
-------signal peptide-----------------------------------------|

       250        260        270        280        290        3002
CTGGCCGACTTCAGCCTGGACAACGAGGTGCACTCGAGCTTCATCCACCGGCGCCTCCGC
 L  A  D  F  S  L  D  N  E  V  H  S  S  F  I  H  R  R  L  R
------|
       310        320        330        340        350        360
AGCCAGGAGCGGCGGGAGATGCAGCGCGAGATCCTCTCCATTTTGGGCTTGCCCCACCGC
 S  Q  E  R  R  E  M  Q  R  E  I  L  S  I  L  G  L  P  H  R
       370        380        390        400        410        420
CCGCGCCCGCACCTCCAGGGCAAGCACAACTCGGCACCCATGTTCATGCTGGACCTGTAC
 P  R  P  H  L  Q  G  K  H  N  S  A  P  M  F  M  L  D  L  Y
       430        440        450        460        470        480
AACGCCCATGGCGGTGGAGGAGGGCGGCGGCCCGGCGGCCAGGGCTTCTCCTACCCCTAC
 N  A  H  G  G  G  G  G  R  R  P  G  G  Q  G  F  S  Y  P  Y
       490        500        510        520        530        540
AAGGCCGTCTTCAGTACCCAGGGCCCCCCTCTGGCCAGCCTGCAAGATAGCCATTTCCTC
 K  A  V  F  S  T  Q  G  P  P  L  A  S  L  Q  D  S  H  F  L
       550        560        570        580        590        600
ACCGACGCCGACATGGTCATGAGCTTCGTCAACCTCGTGGAACATGACAAGGAATTCTTC
 T  D  A  D  M  V  M  S  F  V  N  L  V  E  H  D  K  E  F  F
       610        620        630        640        650        660
CACCCACGCTACCACCATCGAGAGTTCCGGTTTGATCTTTCCAAGATCCCAGAAGGGGAA
 H  P  R  Y  H  H  R  E  F  R  F  D  L  S  K  I  P  E  G  E
       670        680        690        700        710        720
GCTGTCACGGCAGCCGAATTCCGGATCTACAAGGACTACATCCGGGAACGCTTCGACAAT
 A  V  T  A  A  E  F  R  I  Y  K  D  Y  I  R  E  R  F  D  N
       730        740        750        760        770        780
GAGACGTTCCGGATCAGCGTTTATCAGGTGCTCCAGGAGCACTTGGGCAGGGAATCGGAT
 E  T  F  R  I  S  V  Y  Q  V  L  Q  E  H  L  G  R  E  S  D
       790        800        810        820        830        840
CTCTTCCTGCTCGACAGCCGTACCCTCTGGGCCTCGGAGGAGGGCTGGCTGGTGTTTGAC
 L  F  L  L  D  S  R  T  L  W  A  S  E  E  G  W  L  V  F  D
```

```
       850       860       870       880       890       900
ATCACAGCCACCAGCAACCACTGGGTGGTCAATCCGCGGCACAACCTGGGCCTGCAGCTC
 I  T  A  T  S  N  H  W  V  V  N  P  R  H  N  L  G  L  Q  L
       910       920       930       940       950       960
TCGGTGGAGACGCTGGATGGGCAGAGCATCAACCCCAAGTTGGCGGGCCTGATTGGGCGG
 S  V  E  T  L  D  G  Q  S  I  N  P  K  L  A  G  L  I  G  R
       970       980       990      1000      1010      1020
CACGGGCCCCAGAACAAGCAGCCCTTCATGGTGGCTTTCTTCAAGGCCACGGAGGTCCAC
 H  G  P  Q  N  K  Q  P  F  M  V  A  F  F  K  A  T  E  V  H
      1030      1040      1050      1060      1070      1080
TTCCGCAGCATCCGGTCCACGGGGAGCAAACAGCGCAGCCAGAACCGCTCCAAGACGCCC
 F  R  S  I  R  S  T  G  S  K  Q  R  S  Q  N  R  S  K  T  P
                           *  *  *  *  *  *  *  *  *  *  *
      1090      1100      1110      1120      1130      1140
AAGAACCAGGAAGCCCTGCGGATGGCCAACGTGGCAGAGAACAGCAGCAGCGACCAGAGG
 K  N  Q  E  A  L  R  M  A  N  V  A  E  N  S  S  S  D  Q  R
 *
      1150      1160      1170      1180      1190      1200
CAGGCCTGTAAGAAGCACGAGCTGTATGTCAGCTTCCGAGACCTGGGCTGGCAGGACTGG
 Q  A  C  K  K  H  E  L  Y  V  S  F  R  D  L  G  W  Q  D  W
      1210      1220      1230      1240      1250      1260
ATCATCGCGCCTGAAGGCTACGCCGCCTACTACTGTGAGGGGGAGTGTGCCTTCCCTCTG
 I  I  A  P  E  G  Y  A  A  Y  Y  C  E  G  E  C  A  F  P  L
      1270      1280      1290      1300      1310      1320
AACTCCTACATGAACGCCACCAACCACGCCATCGTGCAGACGCTGGTCCACTTCATCAAC
 N  S  Y  M  N  A  T  N  H  A  I  V  Q  T  L  V  H  F  I  N
      1330      1340      1350      1360      1370      1380
CCGGAAACGGTGCCCAAGCCCTGCTGTGCGCCCACGCAGCTCAATGCCATCTCCGTCCTC
 P  E  T  V  P  K  P  C  C  A  P  T  Q  L  N  A  I  S  V  L
      1390      1400      1410      1420      1430      1440
TACTTCGATGACAGCTCCAACGTCATCCTGAAGAAATACAGAAACATGGTGGTCCGGGCC
 Y  F  D  D  S  S  N  V  I  L  K  K  Y  R  N  M  V  V  R  A
      1450      1460      1470      1480      1490      1500
TGTGGCTGCCACTAGCTCCTTCCGAGAATTCCAGACCTTTGGGGCCCAAAGGTTTTTCTG
 C  G  C  H  *
      1510
GATCC
```

OP1D

```
        leader                :  OP1--------->
M K A I F V L K G S L D R D P S S F I H R R L R S Q E R R E
M Q R E I L S I L G L P H R P R P H L Q G K H N S A P M F M
L D L Y N A M A V E E G G G P G G Q G F S Y P Y K A V F S T
Q G P P L A S L Q D S H F L T D A D M V M S F V N L V E H D
K E F F H P R Y H H R E F R F D L S K I P E G E A V T A A E
F R I Y K D Y I R E R F D N E T F R I S V Y Q V L Q E H L G
R E S D L F L L D S R T L W A S E E G W L V F D I T A T S N
H W V V N P R H N L G L Q L S V E T L D G Q S I N P K L A G
L I G R H G P Q N K Q P F M V A F F K A T E V H F R S I R S
T G S K Q R S Q N R S K T P K N Q E A L R M A N V A E N S S
S D Q R Q A C K K H E L Y V S F R D L G W Q D W I I A P E G
Y A A Y Y C E G E C A F P L N S Y M N A T N H A I V Q T L V
H F I N P E T V P K P C C A P T Q L N A I S V L Y F D D S S
N V I L K K Y R N M V V R A C G C H
```

CBMP2B1

```
         10         20         30         40         50         60
ATGAAAGCAATTTTCGTACTGAAAGGTTCACTGGACAGAGATCTGGACTCTCGTCTGGAT
 M  K  A  I  F  V  L  K  G  S  L  D  R  D  L  D  S  R  L  D
         70         80         90        100        110        120
CTGGACGTTCGTACCGACCACAAAGACCTGTCTGATCACCTGGTTCTGGTCGACCTGGCT
 L  D  V  R  T  D  H  K  D  L  S  D  H  L  V  L  V  D  L  A
        130        140        150        160        170        180
CGTAACGACCTGGCTCGTATCGTTACTCCCGGGTCTCGTTACGTTGCGGATCCTAAGCAT
 R  N  D  L  A  R  I  V  T  P  G  S  R  Y  V  A  D::P  K  H
        190        200        210        220        230        240
CACTCACAGCGGGCCAGGAAGAAGAATAAGAACTGCCGGCGCCACTCGCTCTATGTGGAC
 H  S  Q  R  A  R  K  K  N  K  N  C  R  R  H  S  L  Y  V  D
        250        260        270        280        290        300
TTCAGCGATGTGGGCTGGAATGACTGGATTGTGGCCCCACCAGGCTACCAGGCCTTCTAC
 F  S  D  V  G  W  N  D  W  I  V  A  P  P  G  Y  Q  A  F  Y
```

```
        310        320        330        340        350        360
TGCCATGGCGAATGCCCTTTCCCGCTAGCGGATCACTTCAACAGCACCAACCACGCCGTG
 C  H  G  E  C  P  F  P  L  A  D  H  F  N  S  T  N  H  A  V
        370        380        390        400        410        420
GTGCAGACCCTGGTGAACTCTGTCAACTCCAAGATCCCTAAGGCTTGCTGCGTGCCCACC
 V  Q  T  L  V  N  S  V  N  S  K  I  P  K  A  C  C  V  P  T
        430        440        450        460        470        480
GAGCTGTCCGCCATCAGCATGCTGTACCTGGACGAGAATGAGAAGGTGGTGCTGAAGAAC
 E  L  S  A  I  S  M  L  Y  L  D  E  N  E  K  V  V  L  K  N
        490        500        510        520
TACCAGGAGATGGTAGTAGAGGGCTGCGGCTGCCGCTAACTGCAG
 Y  Q  E  M  V  V  E  G  C  G  C  R  *
```

CBMP2B2

```
M K A I F V L K G S L D R D L D S R L D L D V R T D H K D L
S D H L V L V D L A R N D L A R I V T P G S R Y V A D L E F
P G:E L L R D F E A T L L Q M F G L R R R P Q P S K S A V I
P D Y M R D L Y R L Q S G E E E E E Q I H S T G L E Y P E R
P A S R A N T V R S F H H E E H L E N I P G T S E N S A F R
F L F N L S S I P E N E A I S S A E L R L F R E Q V D Q G P
D W E R G F H R I N I Y E V M K P P A E V V P G H L I T R L
L D T R L V H H N V T R W E T F D V S P A V L R W T R E K Q
P N Y G L A I E V T H L H Q T R T H Q G Q H V R I S R S L P
Q G S G N W A Q L R P L L V T F G H D G R G H A L T R R R R
:A K R S P K H H S Q R A R K K N K N C R R H S L Y V D F S D
V G W N D W I V A P P G Y Q A F Y C H G D C P F P L A D H L
N S T N H A I V Q T L V N S V N S S I P K A C C V P T E L S
A I S M L Y L D E Y D K V V L K N Y Q E M V V E G C G C R
```

a longer

a Construct OP1(a) is a cDNA sequence encoding ~~what is believed to be the mature~~ form of OP1 linked by an a Asp-Pro (D-P) dilute acid cleavage site to a leader

sequence ("MLE leader", MKAIF...ADLEF) suitable for promoting ~~expressing~~ expression in E. coli. OP1(b) comprises the MLE leader linked to a portion of the precursor region of the prepro form of OP1 (RIYKD...SKTPK) and the mature form (extending through ACGCH). OP1(c) comprises a leader peptide and the full prepro form of OP1 cDNA including the presumed signal peptide (MHVRS...RSALA). OP1d (amino acid sequence only) comprises a leader sequence ("short TRP," MKA...LDR), an Asp Pro cleavage site, and the presumed entire pro form of the OP1 protein (SSFIH...ACGCH). CBMP2b1 comprises the MLE leader linked through an Asp-Pro dilute acid cleavage site to the presumed mature region of CBMP2b (PKHHS...GCGCR). Approximately one half of this construct comprised cDNA; the other half was synthesized from oligonucleotides. CBMP2b2 comprises the MLE leader linked to the presumed full length pro form of CBMP2b (ELLRD...GCGCR).

The genes were expressed in E. coli under the influence of a synthetic trp promoter-operator to produce insoluble inclusion bodies. The inclusion bodies were solubilized in 8M urea following cell lysis, dialyzed against 1% acetic acid, and partly purified by differential solubilization. Constructs containing the Asp-Pro sites were cleaved with dilute acid. The resulting products were passed through a Sephacryl-200HR or SP Trisacyl column to further purify the proteins, and then subjected to HPLC on a semi-prep C-18 column to separate the leader proteins and other minor impurities from the OP, or BMP2 constructs. Both the BMP2 and OP1 proteins may be purified by chromatography on heparin sepharose. The output of the HPLC column was lyophilized at pH=2 so that it remains reduced.

Conditions for refolding were at pH 8.0 using Tris buffer and 6M glutathione HCl at a protein concentration of several mg/ml. Those solutions were diluted with water to produce a 2M or 3M glutathione concentration and left for 18 hours at 4°C. Air dissolved or entrained in the buffer assures oxidation of the protein in these circumstances.

Samples of the various purified constructs and various mixtures of pairs of the constructs refolded together were applied to SDS polyacrylamide gels, separated by electrophoresis, sliced, incorporated in a matrix as disclosed below, and tested for osteogenic activity. These studies demonstrated that both homodimers and heterodimers of each of the species are active whereas monomers are not.

After N terminal sequencing of the various constructs to confirm their identity, polyclonal antisera against the recombinant presumed mature form proteins were produced. The human OP1 antisera reacted with both the glycosylated and deglycosylated higher molecular weight subunits of naturally sourced bovine material. Antisera against recombinant mature human CBMP2 reacted with both the glycosylated and deglycosylated lower molecular weight subunit of naturally sourced bovine material. While there was some cross-reactivity, this was expected in view of the significant homology between CBMP2 and OP1 (approx. 60%), and the likelihood that degraded OP1 generated during purification contaminates the lower molecular weight subunit. Both antisera react with the naturally sourced 30 Kd dimeric BOP.

Pyridylethylation of C-18 purified, reduced, bovine OP, followed by separation by SDS-PAGE, gel slicing, elution, and digestion with endoproteinase Asp-N, then

Staph V-8 protease, permitted separation of peptide fragments representative of each of the subunits from natural sourced bovine material. Sequencing of the peptide fragments yielded five sequences unequivocally from OP1, six unequivocally from CBMP2(a), and three that could have been from either CBMP2(a) or CBMP2(b). Another peptide, analyzed as D-X-P-F-P-L, was consistent with the presence of CBMP2(b). However, the amino terminal aspartic acid could have been a glutamic acid (E), in which case the peptide would have indicated the presence of CBMP2(a). The DXPFPL sequence determination has not been repeated successfully. From these data, it is apparent that the active osteogenic protein comprises OP1 and CBMP2 only.

The prepro form amino acid sequence of OP1 and the CBMP2 species are set forth in the drawing. Assuming the pro forms of these proteins are cleaved in a way similar to the TGF protein family, and given a molecular weight in the range indicated by the SDS-PAGE separations of deglycosylated, reduced, highly purified material, it is hypothesized that the N termini of the OP1 protein is in or C-terminal to the amino acid residues starred in the drawing. The designation "::" indicates the presumed cleavage sites of CBMP2(a) and (b). The signal peptide of OP1 is proposed as being 29 AA long with two arginine residues in the N-terminal region and one arginine close to the cleavage site. The signal peptide of TGF-ß1 is 29 AA long and that of TGF-ß3 is 18 or 23 AA long. Mature OP1 contains 3 potential N glycosylation sites; there is an additional site in the precursor region. CBMP2 contain 4 or 5 glycosylation sites.

MATRIX PREPARATION

Demineralized, guanidine extracted, particulate allogenic bone may be used as a carrier for osteogenic protein. This type of carrier may be replaced by either a biodegradable-synthetic or synthetic-inorganic matrix (e.g., hydroxyapatite, collagen, tricalcium phosphate, or polylactic acid, polyglycolic acid and various copolymers thereof). Also, xenogenic bone powder may be used if pretreated as described below.

Studies have shown that surface charge, particle size, the presence of mineral, and the methodology for combining matrix and osteogenic protein all play a role in achieving successful bone induction. Perturbation of the charge by chemical modification abolishes the inductive response. Particle size influences the quantitative response of new bone; particles between 75 and 420 mm elicit the maximum response. Contamination of the matrix with bone mineral will inhibit bone formation. Most importantly, the procedures used to formulate osteogenic protein onto the matrix are extremely sensitive to the physical and chemical state of both the osteogenic protein and the matrix.

The sequential cellular reactions at the interface of the bone matrix/OP implants are complex. The multistep cascade includes: binding of fibrin and fibronectin to implanted matrix, chemotaxis of cells, proliferation of fibroblasts, differentiation into chondroblasts, cartilage formation, vascular invasion, bone formation, remodeling, and bone marrow differentiation.

A successful carrier for osteogenic protein must perform several important functions. It must bind osteogenic protein and act as a slow release delivery system, accommodate each step of the cellular response during bone development, and protect the osteogenic protein from nonspecific proteolysis. In addition, selected materials must be biocompatible in vivo and biodegradable; the carrier must act as a temporary scaffold until replaced completely by new bone. In bones, the dissolution rates can vary according to whether the implant is placed in cortical or trabecular bone.

Matrix geometry, particle size, the presence of surface charge, and porosity or the presence of interstices among the particles of a size sufficient to permit cell infiltration, are all important to successful matrix performance. It is preferred to shape the matrix to the desired form of the new bone and to have dimensions which span non-union defects. Rat studies show that the new bone is formed essentially having the dimensions of the device implanted.

The matrix may comprise a shape-retaining solid made of loosely adhered particulate material, e.g., with collagen. It may also comprise a molded, porous solid, or simply an aggregation of close-packed particles held in place by surrounding tissue. Masticated muscle or other tissue may also be used. Large allogenic ~~allogeneic~~ bone implants can act as a carrier for the matrix if their marrow cavities are cleaned and packed with particles and the dispersed osteogenic protein.

Allogenic Matrix

Allogenic demineralized bone matrix is prepared using dehydrated bone as described herein to produce a bone particle size which ~~pass~~ passes through a 420 μm sieve. The bone particles are subjected to dissociative extraction with 4 M guanidine-HCl. Such treatment results in a complete loss of the inherent ability of the bone matrix to induce endochondral bone differentiation. The remaining insoluble material is used to fabricate the matrix. The material is mostly collagenous in nature, and in the absence of pretreatment with OP does not induce cartilage or bone. The total loss of biological activity of bone matrix is restored when an active osteoinductive protein fraction or a pure protein is reconstituted with the biologically inactive insoluble collagenous matrix. Because of the scarcity and high cost of human bone, other matrix options are preferred in clinical settings. However, allogenic bone treated as described ~~are effective as carriers~~ is effective as a carrier for the osteogenic protein.

When osteogenic protein is reconstituted with collagenous bone matrix from other species and implanted in rat, no bone is formed. This suggests that while the osteogenic protein is effective cross species (not species specific), the matrix is species specific, and cannot be implanted cross species. Thus, heretofore, for bone-based matrices, in order for the osteogenic protein to exhibit its full bone inducing activity, a species specific collagenous bone matrix was required.

The major component of all bone matrices is Type I collagen, which is glycosylated. In addition to collagen, extracted bone includes non-collagenous proteins which may

account for 5% of its mass. Many non-collagenous components of bone matrix are glycoproteins. Although the biological significance of the glycoproteins in bone formation is not known, the glycoproteins may present themselves as potent antigens by virtue of their carbohydrate content and may constitute immunogenic and/or inhibitory components that are present in xenogenic matrix.

Xenogenic Matrix

A collagenous bone matrix may be used as a carrier to effect bone inducing activity in xenogenic implants, if one first treats the bone powder as disclosed below.

Bovine bone residue prepared as described above is sieved, and particles of the appropriate size are collected. The sample is dried *in vacuo* over P_2O_5, transferred to the reaction vessel, and exposed to anhydrous hydrogen fluoride (10-20 ml/g of matrix) by distillation onto the sample at -70°C. The vessel is allowed to warm to 0°C and the reaction mixture is stirred at this temperature for 120 min. After evaporation of the HF *in vacuo*, the residue is dried thoroughly *in vacuo* over KOH pellets to remove any remaining traces of acid. Extent of deglycosylation can be determined from carbohydrate analysis of matrix samples taken before and after treatment with HF, after washing the samples appropriately to remove non-covalently bound carbohydrates. SDS extracted protein from HF treated material is negative for carbohydrate as determined by Con A blotting.

The deglycosylated bone matrix is next treated as set forth below:

1) suspend in TBS (Tris-buffered saline) 1g/200 ml, and stir at 4°C for 2 hrs; or in 6 M urea, 50 mM Tris-HCl, 500 mM NaCl, pH 7.0 (UTBS) and stir at room temperature (RT) for 30 min;

2) centrifuge and wash with TBS or UTBS; and

3) centrifuge; discard supernatant; water wash residue; and then lyophilize.

C2. Trifluoroacetic acid

Like hydrogen fluoride, trifluoroacetic acid (TFA) is known to cause swelling of proteins. However, it does not effect deglycosylation.

Bovine bone residue, prepared as described above is sieved, and particles of the appropriate size are extracted with various percentage (1.0% to 100%) of trifluoroacetic acid in water (v/v) at 0°C or RT for 1-2 hours with constant stirring. The treated matrix is filtered, lyophilized or washed with water/salt and then lyophilized.

C3. Dichloromethane

Dichloromethane (DCM) is an organic solvent capable of denaturing proteins without affecting their primary structure. It is a common reagent in automated peptide synthesis, and is used in washing steps to remove unwanted components. DCM does not cause deglycosylation.

Bovine bone residue particles of the appropriate size prepared as described above are incubated for one or two hours at 0°C, and also at RT for the same duration. After the treatment, the matrix is washed with the standard 6M urea containing buffer, or water alone. Alternatively, the matrix is treated with DCM many times (X3) with short washes (20 min. each) with no incubation.

C4. Acetonitrile

Acetonitrile (ACN) is an organic solvent, capable of denaturing proteins without affecting their primary structure. It is a common reagent in high performance liquid chromatography, and is used to elute proteins from silica based column by perturbing hydrophobic interactions. Acetonitrile does not cause deglycosylation.

Bovine bone residue particles of the appropriate size prepared as described above are treated with 100% acetonitrile (1.0g/30ml) at room temperature for one to two hours with constant stirring. The treated matrix is then water washed, or washed with urea buffer, or 4M NaCl, and lyophilized.

C5. Isopropanol

Isopropanol also is an organic solvent capable of denaturing proteins without affecting their primary structure. It is a common reagent used to elute proteins from silica HPLC columns. Isopropanol does not cause deglycosylation.

Bovine bone residue particles of the appropriate size prepared as described above are treated with 100% isopropanol (1.0g/30ml) at room temperature for one to two hours with constant stirring. The treated matrix is then water washed, or washed with urea buffer or 4M NaCl before being lyophilized.

C6. Combinations of Reagents.

Separate bovine bone particle samples are treated with dichloromethane, or acetonitrile, and isopropanol, each of which contained 0.1% trifluoroacetic acid. The optimal conditions for the treatment are incubation with solvent/acid mixture at 0°C or RT for one to two hours with constant stirring. The treated matrix is then lyophilized without wash. Alternately, the treated matrices are washed with water or 4M salt before lyophilization. This type of matrix preparation is currently preferred.

Treatment as set forth above in the swelling agents and other reagents is effective to assure that the material is free of pathogens prior to implantation.

To produce an active osteogenic powder, any of the matrices described above are added to osteogenic protein dissolved in guanidine-HCl. Samples are vortexed and incubated at a low temperature. Cold absolute ethanol is added to the mixture which is then stirred and incubated. After centrifugation (microfuge, high speed) the supernatant is discarded. The matrix is washed with cold concentrated ethanol in water and then lyophilized.

In an alternative procedure, osteogenic protein in an acetonitrile trifluoroacetic acid (ACN/TFA) solution was added to the carrier. Samples were vigorously vortexed many times and then lyophilized.

For those proteins that are prepared in urea buffer, the protein may be mixed with the matrix material, vortexed many times, and then lyophilized. the lyophilized material may be used "as is" for implants.

For the purest, most active OP samples prepared to date, approximately 2-3 ng of proteins per mg of matrix is sufficient to induce bone formation. Depending on the nature and purity of the preparation, dosage may range up to 20 µg/mg implant.

The material is a fine powder, insoluble in water, comprising nonadherent particles. It may be used simply by packing into the volume where new bone growth is desired, held in place by surrounding tissue, and then immobilizing the region to permit osteogenesis. Alternatively, the powder may be encapsulated in, e.g., a gelatin or polylactic acid coating, which is absorbed readily by the body. The powder may be shaped to a volume of given dimensions and held in that shape by interadhering the particles using, for example, soluble, species-biocompatible collagen.

BIOASSAY

The recombinant proteins set forth above and other constructs have been incorporated in allogenic and xenogenic matrices to produce osteogenic devices, and assayed in rat for endochondral bone. No activity is

observed if the matrix is implanted alone. The bioassay for bone induction, as described by Sampath and Reddi (Proc. Natl. Acad. Sci. USA (1983) 80: 6591-6595), is incorporated herein by reference. This assay consists of implanting the test samples in subcutaneous sites in allogeneic recipient rats under ether anesthesia. Male Long-Evans rats, aged 28-32 days, were used. A vertical incision (1 cm) is made under sterile conditions in the skin over the thoraic region, and a pocket is prepared by blunt dissection. Approximately 25 mg of the test sample is implanted deep into the pocket and the incision is closed with a metallic skin clip. The day of implantation is designated as day one of the experiment. Implants were removed on day 12. The heterotropic site allows for the study of bone induction without the possible ambiguities resulting from the use of orthotopic sites.

The implant model in rats exhibits a controlled progression through the stages of matrix induced endochondral bone development including: (1) transient infiltration by polymorphonuclear leukocytes on day one; (2) mesenchymal cell migration and proliferation on days two and three; (3) chondrocyte appearance on days five and six; (4) cartilage matrix formation on day seven; (5) cartiliage calcification on day eight; (6) vascular invasion, appearance of osteoblasts, and formation of new bone on days nine and ten; (7) appearance of osteoblastic and bone remodeling and dissolution of the implanted matrix on days twelve to eighteen; and (8) hematopoietic bone marrow differentiation in the ossicle on day twenty-one. The results show that the shape of the new bone conforms to the shape of the implanted matrix.

Histological sectioning and staining is preferred

to determine the extent of osteogenesis in the implants. Implants are fixed in Bouins Solution, embedded in parafilm, cut into 6-8 μm sections. Staining with toluidine blue or hemotoxylin/eosin demonstrates clearly the ultimate development of endochondrial bone. Twelve day implants are usually sufficient to determine whether the implants show bone inducing activity.

Histological examination of implants indicate homodimers of each of the constructs disclosed above have true osteogenic activity. In addition, osteogenic activity is present with OP1a-CBMP2b1, OP1b-CBMP2b1, and OP1c-CBMP2b2 heterodimer. Homodimeric and heterodimeric structures of the truncated analog active regions set forth below also induce osteogenesis as determined by histological examination.

```
          1       10          20         30         40
COP5           LYVDFS-DVGWDDWIVAPPGYQAFYCHGECPFPLAD
                     50         60         70
          HFNSTN--H-AVVQTLVNSVNSKI--PKACCVPTELSA
              80         90         100
          ISMLYLDENEKVVLKNYQEMVVEGCGCR

          1       10          20         30         40
COP7           LYVDFS-DVGWNDWIVAPPGYHAFYCHGECPFPLAD
                     50         60         70
          HLNSTN--H-AVVQTLVNSVNSKI--PKACCVPTELSA
              80         90         100
          ISMLYLDENEKVVLKNYQEMVVEGCGCR
```

The invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. The present embodiments are therefore to be considered in all respects as illustrative and not restrictive, the scope of the invention being indicated by the appended claims rather than by the foregoing description, and

all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced therein.

What is claimed is: